Arrays and Matrices

Estimated time needed: **20** minutes

Objectives

After completing this lab you will be able to:

- Understand what is an array by coding practices
- Operate on arrays
- Understand what is a matrix by coding practices
- Operate on matrices

Table of Contents

This lesson will cover some basics concepts about vector and factors in R:

About the Dataset

Imagine you got many movie recomendations from your friends and compiled all of the recomendations in a table, with specific info about each movie.

The table has one row for each movie and several columns

- **name** - The name of the movie
- **year** - The year the movie was released
- **length_min** - The lenght of the movie in minutes
- **genre** - The genre of the movie
- **average_rating** - Average rating on Imdb
- **cost_millions** - The movie's production cost in millions
- **sequences** - The amount of sequences
- **foreign** - Indicative of whether the movie is foreign (1) or domestic (0)
- **age_restriction** - The age restriction for the movie

You can see part of the dataset below

name	year	length_min	genre	average_rating	cost_millions	foreign	age_restriction
Toy Story	1995	81	Animation	8.3	30	0	0
Akira	1998	125	Animation	8.1	10.4	1	14
The Breakfast Club	1985	97	Drama	7.9	1	0	14
The Artist	2011	100	Romance	8	15	1	12
Modern Times	1936	87	Comedy	8.6	1.5	0	10
Fight Club	1999	139	Drama	8.9	63	0	18
City of God	2002	130	Crime	8.7	3.3	1	18
The Untouchables	1987	119	Drama	7.9	25	0	14
Star Wars	1977	121	Action	8.7	11	0	10
American Beauty	1999	122	Drama	8.4	15	0	14
Room	2015	118	Drama	8.3	13	1	14
Dr. Strangelove	1964	94	Comedy	8.5	1.8	1	10
The Ring	1998	95	Horror	7.3	1.2	1	18
Monty Python and the Holy Grail	1975	91	Comedy	8.3	0.4	1	18
High School Musical	2006	98	Comedy	5.2	4.2	0	0
Shaun of the Dead	2004	99	Horror	8	6.1	1	18
Taxi Driver	1976	113	Crime	8.3	1.3	1	14
The Shawshank Redemption	1994	142	Crime	9.3	25	0	16
Interstellar	2014	169	Adventure	8.6	165	0	10
Casino	1995	178	Biography	8.2	50	0	18

The Goodfellas	1990	145	Biography	8.7	25	0	14
Blue is the Warmest Colour	2013	179	Romance	7.8	4.5	1	18
Black Swan	2010	108	Thriller	8	13	0	16
Back to the Future	1985	116	Sci-fi	8.5	19	0	0
The Wave	2008	107	Thriller	7.6	5.5	1	16
Whiplash	2014	106	Drama	8.5	3.3	1	12
The Grand Hotel Budapest	2014	100	Crime	8.1	25.5	0	14
Jumanji	1995	104	Fantasy	6.9	65	0	12
The Eternal Sunshine of the Spotless Mind	2004	108	Drama	8.3	20	0	14
Chicago	2002	113	Comedy	7.2	45	0	12

What is an Array?

An array is a data structure that holds values grouped together, like a 2 x 2 table of 2 rows and 2 columns. Arrays can also be **multidimensional**, such as a 2 x 2 x 2 array.

What is the difference between an array and a vector?

Vectors are always one dimensional like a single row of data. On the other hand, an array can be multidimensional (stored as rows and columns). The "dimension" indicates how many rows of data there are.

Let's create a 4 x 3 array (4 rows, 3 columns)

The example below is a vector of 9 movie names, hence the data type is the same for all the elements.

[ ]:
```
#Lets first create a vector of nine movies
movie_vector <- c("Akira", "Toy Story", "Room", "The Wave", "Whiplash",
                  "Star Wars", "The Ring", "The Artist", "Jumanji")
movie_vector
```

To create an array, we can use the **array()** function.

[ ]:
```
movie_array <- array(movie_vector, dim = c(4,3))
movie_array
```

Note that **arrays are created column-wise**. Did you also notice that there were only 9 movie names, but the array was 4 x 3? The original **vector doesn't have enough elements** to fill the entire array (that should have 3 x 4 = 12 elements). So R simply fills rest of the empty values by going back to the beginning of the vector and starting again ("Akira", "Toy story", "Room" in this case).

We also needed to provide `c(4,3)` as a second *argument* to specify the number of rows (4) and columns (3) that we wanted.

[Tip] What is an "argument"? How are "arguments" different from "parameters"?
Arguments and parameters are terms you will hear constantly when talking about "functions". - The "parameters" are the input variables used in a function, like "dim" in the function "array()". - The "arguments" refer to the "values" for those parameters that a function takes as inputs, like "c(4,3)"
We actually don't need to write out the name of the parameter (dim) each time, as in: `array(movie_vector, c(4,3))` As long as we write the arguments out in the correct order, R can interpret the code.
Arguments in a function may sometimes need to be of a "specific type". For more information on each function, you can open up the help file by running the function name with a ? beforehand, as in: `?array`

Array Indexing

Let's look at our array again:

[ ]:
```
movie_array
```

To access an element of an array, we should pass in **[row, column]** as the row and column number of that element.
For example, here we retrieve **Whiplash** from row 1 and column 2:

[ ]:
```
movie_array[1,2] #[row, column]
```

To display all the elements of the first row, we should put 1 in the row and nothing in the column part. Be sure to keep in the comma after the `1`.

[ ]:
```
movie_array[1,]
```

Likewise, you can get the elements by column as below.

[ ]:
```
movie_array[,2]
```

To get the dimension of the array, **dim()** should be used.

[ ]:
```
dim(movie_array)
```

We can also do math on arrays. Let's create an array of the lengths of each of the nine movies used earlier.

[ ]:
```
length_vector <- c(125, 81, 118, 81, 106, 121, 95, 100, 104)
length_array <- array(length_vector, dim = c(3,3))
length_array
```

Let's add 5 to the array, to account for a 5-min bathroom break:

[ ]:
```
length_array + 5
```

Coding Exercise: in the code cell below, create a new length vector with 12 elements and create a 4 x 3 array from it

[ ]:
```
# Write your code below. Don't forget to press Shift+Enter to execute the cell
```

▼ Click here for the solution
```
new_length_vector <- c(125, 81, 118, 81, 106, 121, 95, 100, 104, 85, 64, 83)
new_length_array <- array(new_length_vector, dim = c(4,3))
new_length_array
```

Tip: Performing operations on objects, like adding 5 to an array, does not change the object.To change the object, we would need to assign the new result to itself.

Using Logical Conditions to Subset Arrays

Which movies can I finish watching in two hours? Using a logical condition, we can check which movies are less than 2 hours long.

[ ]:
```
mask_array <- length_array < 120
mask_array
```

Using this array of TRUEs and FALSEs, we can subset the array of movie names:

[ ]:
```
x_vector <- c("Akira", "Toy Story", "Room", "The Wave", "Whiplash",
              "Star Wars", "The Ring", "The Artist", "Jumanji")
x_array <- array(x_vector, dim = c(3,3))

x_array[mask_array]
```

Coding Exercise: in the code cell below, find all movie titles with length less than or equal to 90 minutes

Did you know? IBM Watson Studio lets you build and deploy an AI solution, using the best of open source and IBM software and giving your team a single environment to work in. Learn more here.

[ ]:
```
# Write your code below. Don't forget to press Shift+Enter to execute the cell
```

▼Click here for the solution
```
mask_array <- length_array <= 90
x_array[mask_array]
```

What is a Matrix?

Matrices are a subtype of arrays. A matrix **must** have 2 dimensions, whereas arrays are more flexible and can have, one, two, or more dimensions.

To create a matrix out of a vector , you can use **matrix()**, which takes in an argument for the vector, an argument for the number of rows and another for the number of columns.

[ ]:
```
movie_matrix <- matrix(movie_vector, nrow = 3, ncol = 3)
```

[ ]:
```
movie_matrix
```

Coding Exercise: in the code cell below, create a new length vector with 12 elements and create a 4 x 3 matrix from it

[ ]:
```
# Write your code below. Don't forget to press Shift+Enter to execute the cell
```

▼Click here for the solution
```
new_length_vector <- c(125, 81, 118, 81, 106, 121, 95, 100, 104, 85, 64, 80)
new_movie_matrix <- matrix(new_length_vector, nrow = 4, ncol = 3)
new_movie_matrix
```

Accessing elements of a matrix

As with arrays, you can use **[row, column]** to access elements of a matrix. To retrieve "Akira", you should use [1,1] as it lies in the first row and first column.

[ ]:
```
movie_matrix[1,1]
```

To get data from a certain range, the folowing code can help. This takes the elements from rows 2 to 3, and from columns 1 to 2.

[ ]:
```
movie_matrix[2:3, 1:2]
```

Coding Exercise: in the code cell below, get the second row of the matrix

[ ]:
```
# Write your code below. Don't forget to press Shift+Enter to execute the cell
```

▼Click here for the solution
```
movie_matrix[2,]
```

Coding Exercise: in the code cell below, get the second column of the matrix

[ ]:
```
# Write your code below. Don't forget to press Shift+Enter to execute the cell
```

▼Click here for the solution
```
movie_matrix[,2]
```

Concatenation function

A concatenation function is used to combine two vectors into one vector. It combines values of both vectors.
Lets create a new vector for the upcoming movies as a `upcoming_movie` variable and add them to the `movie_vector` variable to create a `new_vector` of movies.

[ ]:
```
upcoming_movie <- c("Fast and Furious 8", "xXx: Return of Xander Cage", "Suicide Squad")
```

[ ]:
```
new_vector <- c(movie_vector, upcoming_movie)
```

[ ]:
```
new_vector
```

Scaling R with big data
As you learn more about R, if you are interested in exploring platforms that can help you run analyses at scale, you might want to sign up for a free account on IBM Watson Studio, which allows you to run analyses in R with two Spark executors for free.

Authors

Hi! It's Hiten Patel, the author of this notebook. I hope you found arrays and matrices easy to learn! As you start learning the foundations of R, feel free to connect with me if you have any questions.

Other Contributors

Yan Luo

Change Log

Date (YYYY-MM-DD)	Version	Changed By	Change Description
2021-03-03	2.0	Yan	Added coding tasks